DOCUMENT RESUME

ED 106 145 SE 019 167

AUTHOR	Taylor, Susan S.
TITLE	The Effects of Mastery, Adaptive Mastery, and Non-Mastery Models on the Learning of a Mathematical Task.
PUB DATE	Mar 75
NOTE	48p.; Paper presented at the annual meeting of the American Educational Research Association (Washington, D.C., March 30 - April 3, 1975). Best Copy Available
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Achievement; *Basic Skills; *Computer Assisted Instruction; Objectives; *Research; Secondary Education; *Secondary School Mathematics; *Teaching Methods
IDENTIFIERS	*Mastery Learning; Research Reports

ABSTRACT

Three learning models (adaptive mastery, typical mastery, and traditional non-mastery learning models) which employed different criteria for terminating computer-based practice in order to determine mastery or non-mastery of arithmetic skills were compared. The efficiency of two different sequencing arrangements (mixed and clustered) of practice items was also examined. All treatments involved the teaching of basic arithmetic skills to seventh-grade students. The adaptive mastery learning model produced the same high level of performance on both the posttest and a delayed retention test as the other two models, but requires less time, fewer practice items, and minimized overpractice. No significant differences were found between the clustered and mixed item arrangements. (Author/SD)

THE EFFECTS OF MASTERY, ADAPTIVE MASTERY, AND NON-MASTERY MODELS ON THE LEARNING OF A MATHEMATICAL TASK

Susan S. Taylor

Control Data Corporation

Paper presented at the Annual Meeting
of the American Educational Research Association
Washington, D. C., March 1975

C-43 The Mastery Learning Approach + Effects

THE EFFECTS OF MASTERY, ADAPTIVE MASTERY, AND NON-MASTERY MODELS ON THE LEARNING OF A MATHEMATICAL TASK

Susan S. Taylor

Control Data Corporation

ABSTRACT

Three learning models (adaptive mastery, typical mastery, and traditional non-mastery learning models) which employed different criteria for terminating computer-based practice in order to determine mastery or non-mastery of arithmetic skills were compared. The efficiency of two different sequencing arrangements (mixed and clustered) of practice items was also examined.

The adaptive mastery learning model produced the same high level of performance on both the posttest and a delayed retention test as the other two models, but required less time, fewer practice items, and minimized overpractice. No significant differences were found between the clustered and mixed item arrangements.

THE EFFECTS OF MASTERY, ADAPTIVE MASTERY, AND NON-MASTERY MODELS ON THE LEARNING OF A MATHEMATICAL TASK[1]

Susan S. Taylor

Control Data Corporation

Mastery learning strategies have been tried in a number of instruction areas at all levels of education and in all size classes. The results of these studies have yielded considerable evidence that mastery learning procedures work well in enabling about 80% of the students to reach the same high level of performance usually attained by less than 20% of the students under non-mastery conditions (Block, 1971a, 1973; Bloom, 1973).

Although there are many variations, a typical mastery learning model consists of the following steps. First, the subject matter is analyzed to determine its component learning elements. Preferably, behavioral objectives are written for each component element. Thus, what the student is expected to master is clearly stated in measurable terms. Next, some initial instruction and practice are provided covering one or more objectives at a time. Immediately after this initial instruction, a formative evaluation of the student's achievement is conducted. This typically consists of a short diagnostic-progress test designed to measure the student's performance on each

1. The research for this study was supported by Florida State University as a part of the author's doctoral dissertation.

of the objectives. The sole purpose of these tests is to provide feedback to the student and teacher regarding learning deficiencies. Consequently, these tests are not used for grading purposes. On the basis of the results from this formative evaluation, each student is prescribed learning correctives for those objectives indicated. The learning correctives can take many forms such as small group sessions, individual tutoring, or a variety of alternative learning materials. Typically, they are supplementary to the regular instruction. After this procedure has been repeated for an entire unit, chapter, or course, a summative evaluation is conducted. The purpose of this test is to measure the student's final achievement and assign grades according to some pre-established criteria so that the grade indicates achievement and not relative class standing. For large blocks of instruction, a review component is occasionally included.

Although few mastery learning studies report using a computer for components of the instructional model, a number of projects investigating computer-managed instruction (CMI) have employed mastery learning strategies. Similar to the typical mastery learning studies, most of the CMI projects also employed a short test for the formative evaluation component. Rivers (1972), however, investigated a more sophisticated strategy for predicting mastery by employing the computer's unique real-time monitoring capability. A number of performance variables measured during computer-based tutorial instruction were utilized by a regression analysis model to predict each student's final achievement. If a student's predicted achievement was below a predetermined level, he was branched to remedial instruction. If his predicted achievement was satisfactory, he continued through the mainline of instruction. Two other treatment groups received either all of the remedial instruction or none of the remedial instruction. A fourth group was given the option of receiving remedial instruction. Results indicated that the regression and all-remediation groups both had higher posttest performance than the other two groups. Also, there was no significant difference in the amount of time taken by the regression model group when compared with the other three groups. One interesting result was that the group which received no remediation took the most amount of time on the mainline instruction.

Computers provide the capability to measure and predict mastery using rather sophisticated techniques. However, first one must define what is meant by mastery, and then conduct research to investigate efficient means for determining mastery. In any case, the continuous monitoring capability of computers can provide considerable assistance in the administration of mastery learning strategies.

One of the purposes of the present study was to investigate the use of computer monitoring of student performance on practice items to predict mastery of mathematical rules. Since the primary difference between computer-administered practice items and computer-administered test items is the provision of feedback for the practice items, a student's performance on the test items would be similar to his performance on the most recent practice items. Thus, the information needed for formative evaluation purposes which is typically obtained from diagnostic progress tests can also be obtained from the most recent practice items. However, appropriate performance measures and criteria for predicting mastery from responses to practice items need to be identified and investigated.

The results of research with respect to the effects of varying the amount of practice on the learning and retention of rules is quite consistent. Additional practice beyond initial learning does not appear to aid retention (Gagne, Mayor, Garstens, and Paradise, 1962; Gay, 1969; Gibson, 1969; Bassler, Curry, Hull, and Mealy, 1971; Hannum, 1973). Therefore, when helping the student to reach mastery, it is critical to provide enough practice to be assured that he has attained the criterion, but it is apparently wasteful to continue to provide examples or practice items beyond the point of reaching criterion.

One study (Gay, 1971) required students to make two consecutive correct answers before proceeding; however, a number of students were unable to exhibit the same level of mastery on a delayed retention test. The idea of using a specified number of consecutive correct answers is appealing because only the student's most recent performance is considered. The question then becomes one of determining how many consecutive correct answers are necessary to determine that mastery has been attained.

Some research by Millman (1972) on determining test length may also apply in this situation, since a criterion of n consecutive correct answers is analogous to a test of n items for which 100% accuracy is required. Millman provides tables of estimated true scores of students passing such tests of varying lengths. Since the ultimate goal of mastery standards is to predict whether or not a student would achieve a specified level of performance on a posttest, then the number of consecutive correct answers should be similar to the length of the test for which a minimum number of students passing the test would have an estimated true score less than the desired level. To minimize the number of students achieving a posttest score of less than 80% correct, a criterion of five consecutive correct answers is suggested. Research is needed to investigate this hypothesis. Also, studies which compare the effectiveness of varying numbers of consecutive correct answers in predicting the desired posttest and retention test performance should be conducted. In view of the results of studies comparing the effects of varying amounts of practice, the number of consecutive correct answers necessary to predict mastery will probably be very small.

Learning Models

For the purpose of investigating the use of student performance on practice items to predict mastery, three learning models were defined. These were a typical mastery learning model, an adaptive mastery learning model, and a traditional, non-mastery learning model. Each model is described below and in Figure 1.

The typical mastery learning model provides some initial instruction followed by a fixed amount of practice. Formative evaluation of student performance is measured by means of a short objective-referenced test. On the basis of the results from this formative evaluation, each student is prescribed learning correctives for those objectives indicated. The learning correctives consist of additional practice with instructional feedback. The formative evaluation and learning corrective steps are repeated for each objective until a specified criterion is achieved on the test. Then a summative evaluation of student performance is conducted.

The adaptive mastery learning model also provides some initial instruction followed by practice. The amount of practice, however, is varied according to each student's performance on the practice items. The formative evaluation of a student's performance for an objective is based on the student's responses to the practice items for that objective. The learning correctives consist of instructional feedback related to the practice items. Practice items are presented until a specified number of consecutive correct responses are made. When this criterion has been met for each objective, a summative evaluation is conducted.

In the traditional, non-mastery learning model, a student is given some initial instruction followed by a fixed number of practice items. Eventually, a summative evaluation is conducted. There is no formative evaluation component, and no attempt is made to correct learning deficiencies.

The major questions explored in comparing the learning models were:

1. Performance: Did the mastery groups achieve a higher level of performance than the non-mastery group?

2. Time: Was there a difference in the practice time required for each model?

3. Amount of practice: Did the adaptive mastery learning model require fewer practice items than the typical mastery learning model?

I. Adaptive Mastery Model	II. Typical Mastery Model	III. Traditional Non-Mastery Model
1. Instruction	1. Instruction	1. Instruction
2. Variable amount of practice until students reach criterion with remedial feedback provided for incorrect responses. This combines practice, formative evaluation, and remedial instruction.	2. Fixed amount of practice (with feedback)	2. Fixed amount of practice (with feedback)
	3. Formative evaluation of student performance	
	4. Remedial instruction if indicated in step 3.	
3. Posttest	5. Posttest	3. Posttest

Figure 1.--Components of Each Model

4. Aptitude: What were the relationships between apptitude and the variables of practice time, performance, and number of practice items presented? Did these conform to Carroll's (1963) predictions?

5. Difficulty: Was there a relationship between the difficulty of the objective and the variables of practice time, performance, and number of practice items presented?

Distribution of Practice

Another issue investigated in this study was that of how the practice items should be arranged when there are several objectives involved. Considerable research has been conducted regarding the distribution of practice over time, but very little attention has been given to the problem of how to present practice items from several different categories within one practice session. The major dependent variables of interest in such research would be the rate of acquisition and amount of retention. Also of importance would be the finding of an interaction between temporal position and type of practice item arrangement.

Johnson (1970) investigated various distributions of related frames within programmed instruction materials used to train Navy technicians. In all cases, both the order in which topics were introduced and the original order of frames within a topic were preserved. The only differences were in the segregation or integration of frames from different topics. In the first study, the group using the more integrated program had significantly higher scores on an immediate posttest. In the second two studies, there were no differences on the immediate posttest, but the more integrated program resulted in significantly higher scores on a retention test administered one week after the completion of the programs. Also, in the third study, the more integrated program format required significantly more reading time.

Reynolds and Glaser (1964) compared linear and spiral versions of a programmed instruction text. There were no significant differences in either the posttest or retention test scores, but the spiral version required 6.3 weeks longer to complete.

Crothers (1965) employed mathematical learning theory to investigate the efficiency of massed versus distributed presentation orders for items from different categories. The massed schedule was predicted to have a higher proportion correct on the test than the distributed schedule when there is little positive transfer. Conversely, when there is a high degree of learning, the distributed schedule is predicted to be superior to the massed schedule. Empirical research is needed to validate these predictions for several types of learning and tasks with varying difficulty.

In summary, the research investigating the efficiency of various presentation orders for items from different categories is sparce and inconclusive. The present study sought to compare two sequencing arrangements. In the clustered arrangement, the students received all practice items relating to one objective consecutively. This procedure was then repeated for all the other objectives. In the mixed item arrangement, a student received all the practice items for several objectives mixed together.

The specific questions explored regarding distribution of practice were:

1. Performance: Did short-term and long-term retention test scores for the two groups differ? Was there a difference in the number of errors made on the practice items themselves?

2. Time: Did the mixed item arrangement require more time than the clustered arrangement?

3. Learning rate: Did the clustered item arrangement require fewer items to reach mastery than the mixed item arrangement?

Method

Subjects

A group of 51 seventh grade students (24 girls and 27 boys) from the Florida State University Developmental Research School participated in the study. All students had the same math teacher. Approximately half of the students had been participating in a self-paced mastery learning project in seventh grade math. The other half had been receiving instruction on the same content in a traditional group-based teacher-oriented setting. Eight additional students who had answered 80% or more of the pretest items correctly were excluded from the study.

Materials

The learning tasks selected for this study dealt with addition, subtraction, multiplication, division, and comparison of integers. Instruction was also provided for a prerequisite objective dealing with the concept f directed numbers, although it was not of central interest in the study. The materials for the study included an instructional workbook for each objective, computer-administered practice sessions, and four criterion tests also presented by computer.

The instructional workbooks were derived from the Student Centered Instructional System (SCIS) math materials developed at the Developmental Research School, Florida State University. Each workbook used a programmed instruction format, but did not include a large block of practice items for each objective.

For each objective, there was also a set of practice problems implemented via computer-assisted instruction (CAI). If the student responded incorrectly to a practice item, brief corrective feedback was provided, and the student was required to respond again. If this second response was also wrong, the correct answer was given along with an explanation. The student was then required to type the correct answer before continuing to the next practice problem. When the student answered a problem correctly, he was given a message such as "Fine!", "Good!", or "Correct!", and then was automatically advanced to the next item after a brief pause. All practice items were in a constructed response format. Item forms and a random number generator were used to create the items. Four item forms representing the four combinations of two positive or negative signs were used for each objective. Two randomly selected numbers ranging from one to nine were inserted in each item form to build a practice item. Within an objective the four item forms were used in rotation such that every fourth item for an objective used the same item form.

All test items were of a constructed response format. All students received exactly the same test items, i.e., item forms and a random number generator were not used for constructing the tests. There were eight items per objective for each test. No feedback or knowledge of results was provided for the test items. The students were, however, told their final score on each test and the final test score was averaged in with their scores on other math tests for that grading period. For all tests, the items were presented using the mixed item arrangement previously described.

The entire study took place at the Computer Applications Laboratory (CAL), Florida State University. All practice items and tests were presented on cathode ray tube (CRT) terminals by an IBM 1500/1800 instructional system. The students completed the workbooks in an adjacent room.

The students' mathematical aptitude was measured by the Mathematics Aptitude Test (R-1) and the Number Facility Tests from the Educational Testing Service Tests for cognitive factors (French, Ekstrom, and Price, 1963). These tests were administered prior to the study. In addition, reading scores from the Comprehensive Tests of Basic Skills were obtained from the Developmental Research School.

Treatments

All subjects were stratified on the basis of total score on the Number Facility Tests and classroom instructional setting and randomly assigned to one of six treatment groups in a 2 x 3 factorial design. The levels of each factor are described below.

Clustered condition. In the clustered practice condition, the subjects were presented with all practice problems related to the first objective. Then all practice problems related to the second objective were presented. This procedure was repeated until the practice problems for all five objectives had been presented as shown in Figure 2.

Mixed condition. For the mixed condition, the practice items for all objectives were presented in rotation such that items relating to a given objective appeared on every fifth presentation (see Figure 2).

Learning model conditions. The treatment groups for each of the three learning models followed the steps outlined in Figure 1. All treatment groups were presented with the same initial instruction in the form of workbooks. The number of practice items presented for each objective depended upon which learning model was employed for each treatment group. For the adaptive mastery learning condition, items for a given objective were presented until the student made five consecutive correct responses on the first attempt. The selection of this criterion was based on previously described research by Millman (1972). When this criterion was met for an objective under the clustered practice condition, the next objective was introduced. As objectives were mastered under the mixed practice condition, the number of items intervening between successive presentations of related items decreased to zero.

For the traditional, non-mastery learning condition, the students were presented with a fixed number of practice items. The number of items presented for each objective was the same as the number of practice problems included in the original SCIS materials. For each objective, the number of practice problems was as follows: Addition - 20; Order - 18; Subtraction - 18; Multiplication - 21; and Division - 14.

For the typical mastery learning condition, the students initially received the same number of practice items per objective as in the

Clustered	Mixed
A1*	A1
A2	B1
A3	C1
A4	D1
.	E1
.	
B1	A2
B2	B2
B3	C2
B4	D2
.	E2
.	
C1	A3
C2	B3
C3	C3
C4	D3
.	E3
.	
D1	A4
D2	B4
D3	C4
D4	D4
.	E4
.	
E1	A5
E2	B5
E3	C5
E4	D5
.	E5
.	.
	.

* A, B, C, D, and E represent the five objectives. 1, 2, 3,..... represent the practice items.

Figure 2.--Order of Presenting Practice Problems Under Clustered and Mixed Conditions

traditional, non-mastery learning condition. However, the students were then administered a diagnostic progress test containing eight items per objective. If the student's score for an objective was less than 80% correct, he was given another set of practice problems related to that objective (equal in number to the original set), followed by a repetition of the associated test items. This practice-testing cycle was repeated until the student reached a criterion of 80% correct on the test items for each objective.

Procedures

All students were brought to the Computer Applications Laboratory to participate in the study. After some preliminary instruction on operating the CRT terminals and the role this study played in their regular math program, a quiz on directed numbers was administered, and remedial instruction provided when necessary. Although the objective on directed numbers was not of central interest in this study, it was considered prerequisite to the subsequent objectives and, consequently, it was necessary to assure that all students possessed the related skills before proceeding.

All students were then administered a forty-item pretest (Test 1) covering addition, order, subtraction, multiplication, and division of integers. The eight students who passed the pretest with a score of 80% or better were automatically excluded from the study. All other students (51) were then given workbooks for each of the five objectives in order. After finishing the last workbook, the students were given a short review consisting of the presentation of rules and examples for all five objectives. Another forty-item test (Test 2) was then administered to measure the students' performance after completing the workbooks and before receiving practice. Following this test, the students were given practice problems according to their treatment condition. Immediately after completing the practice session, each student was given a forty-item posttest (Test 3). The student was then scheduled to return for a retention test (Test 4) one week later.

The time required to complete the initial learning and practice sessions varied from student to student, but took approximately five to 14 days with the students spending 45 minutes each day at CAL. All instruction and testing was conducted on an individually-paced basis.

RESULTS

The data for each measure were analyzed separately by objective as well as by the total scores summed across all five objectives. In most cases the results of the analyses for each objective were consistent with the results obtained for the total scores. Consequently, unless discrepancies were found amongst the objectives, only the analyses of the total scores for each measure are reported here.[2] Since there were no significant interaction effects, only the one way ANOVAs conducted for each factor are reported.

Summary statistics for the aptitude tests and criterion tests are presented in Tables 1 and 2 respectively. The measures taken prior to the treatments include the six aptitude tests, frequency of passes through the directed number quizzes, time spent on the directed number quizzes, performance on Test 1 (Pretest), time to complete the six

2. The complete analyses by objective may be obtained by writing to the author.

TABLE 1

Descriptive Statistics of Aptitude Measures (n=51)

Measure	Mean	Standard Deviation	Reliability Coefficient
Number Facility Tests			
Addition	23.25	7.65	.93
Subtraction & Division	13.82	7.23	.88
Multiplication	36.71	10.83	.88
Mathematical Aptitude*	33.22	12.61	.49
Comprehensive Tests of Basic Skills-Reading (CTBS)			
Vocabulary	25.08	7.58	.79**
Comprehension	28.63	6.45	.71**

*Data transformed ($x' = 4x + 10$)

**KR21 estimate based on the 51 students in this study.

workbooks, time for the review, and performance on Test 2 which was given just before the practice sessions. The analyses of these measures are intended to show the equivalency of the six groups prior to the treatments. A series of 3 x 2 ANOVAs was conducted with learning models and distribution of practice as the independent variables and each of the measures taken prior to the treatment as the dependent variable. No significant differences were found for any of the analyses conducted on measures taken prior to the treatments.

TABLE 2

Descriptive Statistics of Criterion Tests

Measure	Sample Size	Highest Possible Score	Mean	Standard Deviation	Reliability Coefficient
Test 1 (Pretest)					
Addition	51	8	5.4706	2.2392	.68
Order	51	8	5.6863	2.1118	.71
Subtraction	51	8	2.5882	1.4167	.85
Multiplication	51	8	3.0000	1.6125	.79
Division	51	8	3.2157	1.5141	.84
Total	51	40	19.9608	5.3850	.91
Test 2 (Test after workbooks)					
Addition	51	8	5.8627	1.8549	.66
Order	51	8	6.5686	1.8682	.80
Subtraction	51	8	3.4510	2.1290	.75
Multiplication	51	8	5.9020	1.7118	.69
Division	51	8	6.3137	2.1400	.78
Total	51	40	28.0980	6.9922	.91
Test 3 (Posttest)					
Addition	51	8	6.5294	1.9220	.72
Order	51	8	7.3725	1.2955	.76
Subtraction	51	8	6.4706	2.0333	.81
Multiplication	51	8	7.3333	1.2111	.56
Division	51	8	7.3333	1.2437	.64
Total	51	40	35.0392	5.1223	.85
Test 4 (Retention Test after one week)					
Addition	51	8	6.7255	1.3127	.48
Order	51	8	7.3137	1.7028	.88
Subtraction	51	8	6.0588	2.1671	.80
Multiplication	51	8	6.5882	1.0803	.26
Division	51	8	7.3137	1.4070	.72
Total	51	40	34.0000	5.0557	.81

Analyses Relating to Learning Model Hypotheses

The results of the analyses regarding the three learning model treatment groups are presented in the following order: performance measures, time in practice, amount of practice, difficulty of the objectives, and the relationship between aptitude measures and selected dependent variables.

Table 3 presents the means and standard deviations for the three learning model treatment groups of the total scores on the posttest and retention test. A one way ANOVA with learning model as the independent variable and total test score as the dependent variable was conducted for each test. The results of these analyses are reported in Table 4. There were no significant differences on the total scores for either the posttest or the retention test.

TABLE 3

Means and Standard Deviations of Total Posttest and Retention Test Scores for Learning Model Treatment Groups

Measure \ Group	Adaptive Mastery (n=17)	Typical Mastery (n=16)	Traditional Non-Mastery (n=18)
Posttest (Test 3)			
Mean	34.71	36.75	33.83
SD	5.38	3.32	5.98
Retention Test (Test 4)			
Mean	33.94	35.31	32.89
SD	4.97	2.68	6.53

TABLE 4

One Way ANOVAs, Learning Models, for Total Posttest and Retention Test Scores

Source	df	MS	F
Posttest			
Between groups	2	37.44	1.45
Within groups	48	25.77	
Retention Test			
Between groups	2	24.92	.97
Within groups	48	25.59	

An additional performance measure considered was the proportion of students in each treatment group who achieved a score of 80% or better on each test. The observed cell frequencies and the chi-square values obtained for the total posttest and total retention test scores are reported in Table 5. Neither of the chi-square values was significant.

TABLE 5

Chi-Square Analyses of Total Posttest and Retention Test Scores for Learning Model Treatment Groups

% Correct	Learning Model			χ^2
	Adaptive Mastery	Typical Mastery	Traditional Non-Mastery	(df=2)
Posttest Total				
80-100%	14	14	12	3.8781
0-79%	3	2	6	
Retention Test Total				
80-100%	11	15	12	4.5619
0-79%	6	1	6	

The means and standard deviations of the total amount of time spent in practice sessions for each of the three learning model treatment groups are presented in Table 6. A logarithmic transformation was performed on the time measures to normalize the distributions. A one way ANOVA with learning model as the independent variable and total practice time as the dependent variable was conducted using the transformed data. The results of this analysis are reported in Table 7. A substantial F ratio of 16.44 ($p < .001$) was obtained. A Duncan Multiple Range Test yielded significant differences at the .01 level between all pairs of means. The adaptive mastery model required the least amount of practice time and the typical mastery model required the most practice time.

TABLE 6

Means and Standard Deviations of Total Practice Time (in seconds) for Learning Model Treatment Groups

Group / Measure	Adaptive Mastery (n=17)	Typical Mastery (n=16)	Traditional Non-Mastery (n=18)
Mean	1414.35	3757.83	1957.18
SD	1168.87	1742.16	742.38

TABLE 7

One Way ANOVA, Learning Models, for Total Practice Time

Source	df	MS	F
Between groups	2	1.0326	16.44****
Within groups	48	.0628	

****$p < .001$

The number of practice items presented was a variable for the two mastery learning treatment groups, but a constant, by definition, for the traditional non-mastery treatment group. The number of practice items for the five objectives under the traditional non-mastery model was set at 20, 18, 18, 21, and 14, respectively, resulting in a total of 91 practice items. The means and standard deviations of the total number of practice items received for each learning model treatment group are presented in Table 8. A one way ANOVA with the two mastery learning models only as levels of the independent variable and total number of practice items received as the dependent variable was conducted. The results of this analysis are reported in Table 9. An F ratio of 38.71 ($p < .001$) was obtained indicating that the adaptive mastery learning model required significantly fewer items than the typical mastery learning model.

TABLE 8

Means and Standard Deviations of the Total Number of Practice Items for Learning Model Treatment Groups

Group / Measure	Adaptive Mastery (n=17)	Typical Mastery (n=16)	Traditional Non-Mastery (n=18)
Mean	51.47	183.31	91.00*
SD	34.33	79.94	-

*Constant

TABLE 9

One Way ANOVA, Mastery Learning Models
Total Practice Items

Source	df	MS	F
Between groups	1	143,272.21	38.71 ****
Within groups	31	3,700.89	

****$p < .001$

The means of the total number of practice items required for each of the two mastery learning models were compared by means of one-tailed t-tests with 91, the constant number of practice items required by the traditional non-mastery model. Again the adaptive mastery learning model required significantly fewer items ($t = 4.75$, $p < .0005$) than the traditional model while the typical mastery learning model required significantly more practice items ($t = 4.62$, $p < .0005$).

An additional measure taken which relates to the amount of practice required by each model is the length of the longest string of consecutive correct answers per objective which each student made. By definition, the length of the longest string of consecutive correct answers for the adaptive mastery learning model was five for every objective. The means and standard deviations of the average across objectives of this measure for the other two learning models are presented in Table 10.

TABLE 10

Means and Standard Deviations of the Average Length of the Longest String of Consecutive Correct Answers for Learning Model Treatment Groups

Measure \ Group	Adaptive Mastery (n=17)	Typical Mastery (n=16)	Traditional Non-Mastery (n=18)
Mean	5.00*	14.71	11.40
SD	-	3.25	3.30

*Constant

The means of the average length of the longest strings of consecutive correct answers per objective for the typical mastery and traditional, non-mastery learning models were compared by means of one-tailed t-tests with the constant five. The results of these analyses indicated that students in the typical mastery learning model ($t = 11.95$, $p < .0005$) and traditional, non-mastery learning models ($t = 8.23$, $p < .0005$) made significantly more than five consecutive correct responses. Similar comparisons were made for each objective. With one exception, the observed t values were greater than four ($p < .0005$), indicating that the students in these treatment groups made significantly more than five responses. For the traditional, non-mastery group a t value of 0.98 ($p < .20$) was obtained for the subtraction objective.

Another question investigated was the relationship between the difficulty of each objective and the amount of practice required by each of the learning models. The four criterion tests were used to determine the relative difficulty of the five objectives. For each test, the means of the subtests for each objective were ordered from low to high and the ranks from one to five assigned to the corresponding objectives. The Kendall coefficient of concordance W was used to measure the degree of association among the four rankings obtained from the four tests. A W of .849 ($p < .01$) was obtained, indicating strong agreement among the four rankings. Consequently, the four rankings were pooled, which resulted in the following order of objectives from the most difficult to easiest: Objective 4 - Subtraction, Objective 5 - Multiplication, Objective 2 - Addition, Objective 6 - Division, and Objective 3 - Order. This last ranking was used as the standard of difficulty for further comparisons.

Spearman rank correlation coefficients were computed to measure the association between the relative difficulty of the objectives and the relative number of practice items required for each objective by the three learning models. For the adaptive mastery model an r_s of .90 ($p < .05$) was obtained, indicating a significant relationship between the difficulty of the objective and the number of practice items required by the model. The relationships for the typical mastery model ($r_s = .80$) and the traditional non-mastery model ($r_s = .43$) were not significant.

The final question investigated with respect to the three learning model treatments concerned the nature of the relationships between the six aptitude measures and the variables of performance, practice time, and the number of practice items received. The correlations between the aptitude measures and the four dependent variables for each treatment group are presented in Table 11. It was predicted that the correlation

TABLE 11

Correlations Between Aptitude Measures and Various Dependent Variables for Learning Model Treatment Groups

Aptitude Measure	Total Posttest Score	Total Retention Test Score	Practice Time	Number of Practice Items
Adaptive Mastery Group (n=17)				
Number Facility Tests				
Addition	.247	.277	-.403	-.243
Subtraction & Division	.387	.317	-.371	-.312
Multiplication	.353	.293	-.192	-.206
Mathematical Aptitude	.322	.494*	-.522*	-.383
CTBS				
Vocabulary	-.264	-.156	-.025	.082
Comprehension	-.047	.102	-.228	-.065
Typical Mastery Group (n=16)				
Number Facility Tests				
Addition	.318	-.373	.016	-.117
Subtraction & Division	.690**	.437*	-.867**	-.315
Multiplication	.567*	-.027	-.570*	-.421
Mathematical Aptitude	.339	.473*	-.264	.151
CTBS				
Vocabulary	.402	.219	-.505*	-.017
Comprehension	.429*	.082	-.377	.111
Traditional Non-Mastery Group (n=18)				
Number Facility Tests				
Addition	.111	.261	-.274	-
Subtraction & Division	.050	.208	-.072	-
Multiplication	-.014	.138	-.413*	-
Mathematical Aptitude	.220	.149	.047	-
CTBS				
Vocabulary	-.042	-.086	.084	-
Comprehension	.209	.016	-.102	-

* $p < .05$

** $p < .01$

between aptitude and performance would be low under mastery learning conditions and high under traditional non-mastery learning conditions. Quite the opposite results were found in this study in that the highest correlations obtained between aptitude and performance were for the typical mastery learning group with five of the 12 correlations significantly different from zero. The lowest correlations obtained were for the traditional non-mastery group, and none was significant. For the adaptive mastery learning group, only one significant correlation ($p < .05$) was found.

It was also predicted that there would be a large negative correlation between aptitude and time for the mastery learning groups and that this correlation would be near zero for the traditional non-mastery learning group. The results in column three of Table 11 indicate that three of the six correlations for the typical mastery group were significantly different from zero. One correlation for each of the adaptive mastery and traditional non-mastery groups was significant.

Similarly, it was predicted that there would be a large negative correlation between aptitude and number of practice items for the mastery learning groups. By definition, the correlation between aptitude and number of practice items for the traditional non-mastery group is zero, since the number of practice items presented was a constant. There were no significant correlations for the other two treatment groups.

Analyses Relating to Distribution of Practice Hypotheses

The analyses relating to the hypotheses regarding the clustered and mixed arrangements of practice items will be presented in the following order: performance measures, time in practice, amount of practice, and accuracy of responses.

The means and standard deviations of the two treatment groups on the posttest and retention test are presented in Table 12. A on way ANOVA with practice arrangement as the independent variable and total test score as the dependent variable was conducted for each test. The results of these analyses are reported in Table 13. No significant differences were found.

The total amount of practice time as well as the time spent on practice items for each objective was measured. The means and standard deviations of the total amount of time spent in practice sessions for the two distribution treatment groups are shown in Table 14. A logarithmic transformation was performed on the time measures to normalize the distributions. A one way ANOVA with practice arrangement as the independent variable and total practice time as the dependent variable was conducted using the transformed data. The results of this analysis are reported in Table 15. There was no significant difference.

The number of practice items presented was a variable for some of the students. The means and standard deviations of the total number of practice items for the two practice arrangements are presented in Table 16. The data for students who received a fixed number of practice items are not included. A one way ANOVA with practice arrangement as

TABLE 12

Means and Standard Deviations of Total Posttest and Retention Test Scores for Distribution Treatment Groups

Measure \ Group	Clustered (n=25)	Mixed (n=26)
Posttest		
Mean	34.56	35.50
SD	5.20	5.10
Retention Test		
Mean	33.32	34.65
SD	5.76	4.28

TABLE 13

One Way ANOVAs, Distribution, for Total Posttest and Retention Test Scores

Source	df	MS	F
Posttest			
Between groups	1	11.26	.42
Within groups	49	26.54	
Retention Test			
Between groups	1	22.68	.89
Within groups	49	25.62	

TABLE 14

Means and Standard Deviations of Total Practice Time (in seconds) for Distribution Treatment Groups

Measure \ Group	Clustered (n=25)	Mixed (n=26)
Mean	2349.55	2333.06
SD	1762.53	1436.17

TABLE 15

One Way ANOVA, Distribution, for Total Practice Time

Source	df	MS	F
Between groups	1	.0073	.07
Within groups	49	.1035	

TABLE 16

Means and Standard Deviations of the Total Number of Practice Items for Distribution Treatment Groups

Measure \ Group	Clustered (n=16)	Mixed (n=17)
Mean	123.69	107.59
SD	105.38	74.70

the independent variable and the total number of practice items received as the dependent variable was conducted. The results of this analysis are reported in Table 17. No significant difference was found.

TABLE 17

One Way ANOVA, Distribution, Total Practice Items

Source	df	MS	F
Between groups	1	2136.32	.26
Within groups	31	8253.66	

The ratio of the number of correct answers to the total number of responses per objective was calculated for each student as an indicator of the number of errors made. This ratio was necessary to adjust for the varying number of practice items received as well as the fact that multiple responses to the same item were possible. The means and standard deviations of this ratio for all responses are presented in Table 18. A one way ANOVA with practice arrangement as the independent variable and the ratio of correct answers to total responses as the dependent variable was conducted. The results of this analysis are reported in Table 19. No significant difference was found.

Another question investigated concerned the rate of learning for the mixed and clustered item arrangements. For this purpose, a score was calculated for each student indicating the number of errors on the first attempt for the first practice item from each of the five objectives.

TABLE 18

Means and Standard Deviations of the Ratio of the Total Number of Correct Answers to the Total Number of Responses for Distribution Groups

Measure \ Group	Clustered (n=25)	Mixed (n=26)
Mean	.7976	.7897
SD	.1126	.1283

TABLE 19

One Way ANOVA, Distribution, for Ratio of Total Number of Correct Answers to Total Number of Responses

Source	df	MS	F
Between groups	1	.0008	.05
Within groups	49	.0146	

Similar scores were obtained for the second, third, fourth item, and so on. Using these scores, learning curves for each group were plotted and are shown in Figure 3. The X-axis represents the sequential order of item presentations within each objective and the Y-axis represents the mean number of errors within each group. Since the number of items presented to a student varied for each objective, the number of objectives included decreased as the presentation number increased. Consequently, the mean number of errors rather than the mean number

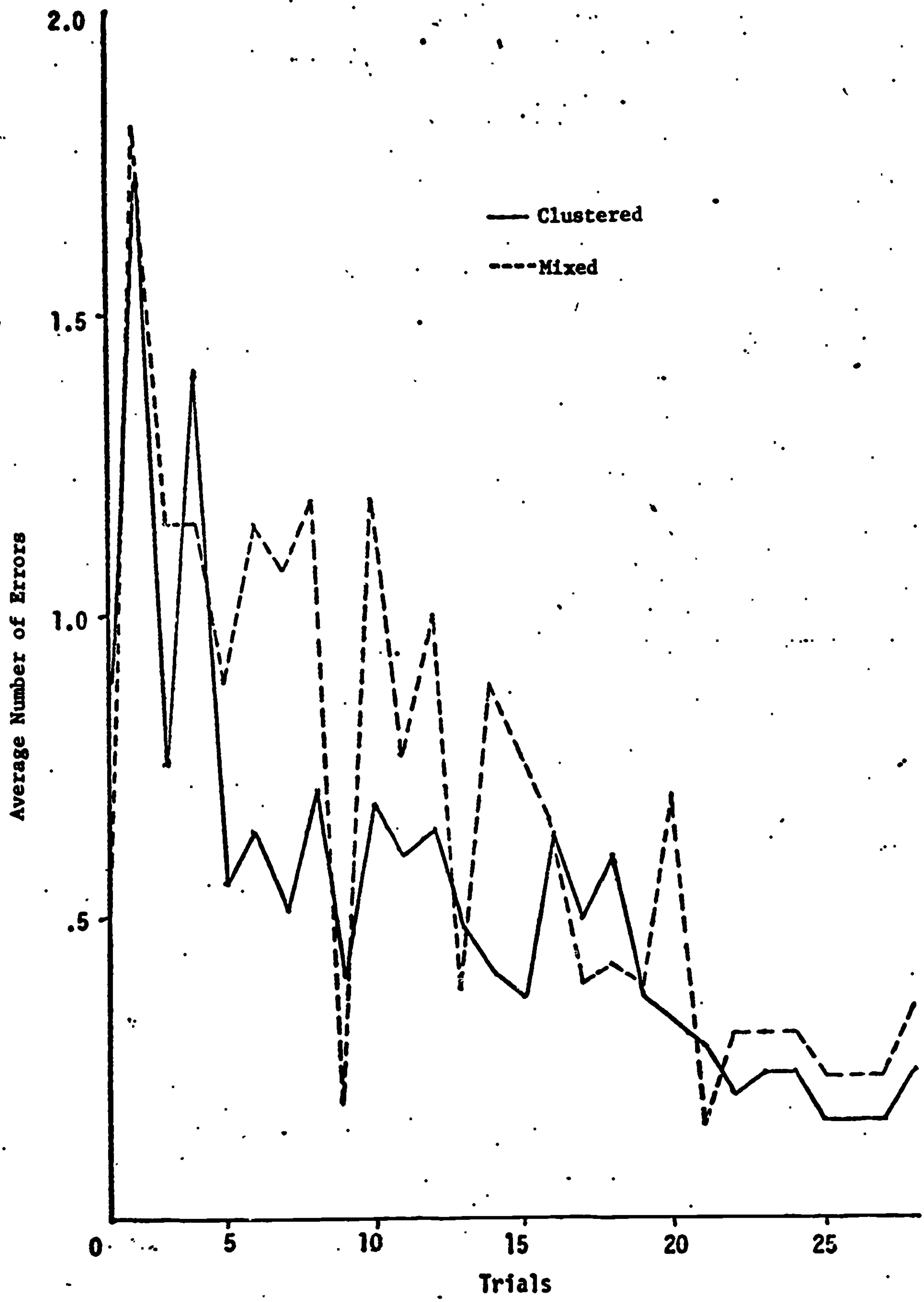

Figure 3. Rate of Learning Curves for the Mixed and Clustered Item Arrangements

of correct responses was plotted. Since both curves dipped on every fourth trial, the means for each block of four trials were calculated and are plotted in Figure 4.

DISCUSSION

The present study sought to investigate two major issues. The first issue concerned the use of students' performance on practice items to determine mastery or non-mastery. Three learning models, which differed in their use of practice item data, were employed to investigate this problem. The second issue was concerned with the effectiveness of mixed and clustered item arrangements for practice.

Learning Models

On the basis of the analyses of Test 2, which was given immediately prior to the practice session, it was established that the three learning model treatment groups were equivalent with respect to learning, prior to the practice sessions.

Performance. There were no significant differences among the three learning model treatment groups with respect to their scores on the posttest and the retention test. Neither were there significant differences in the proportion of students achieving a score of 80% correct on the posttest and retention test for the three learning model groups. The level of performance for all three groups was quite high. The mean scores for all groups on the posttest and retention test exceeded the criterion of 80% correct and a large proportion (75%) of the students achieved mastery on both tests. Thus, it may be concluded that all three models produced the same high level of performance.

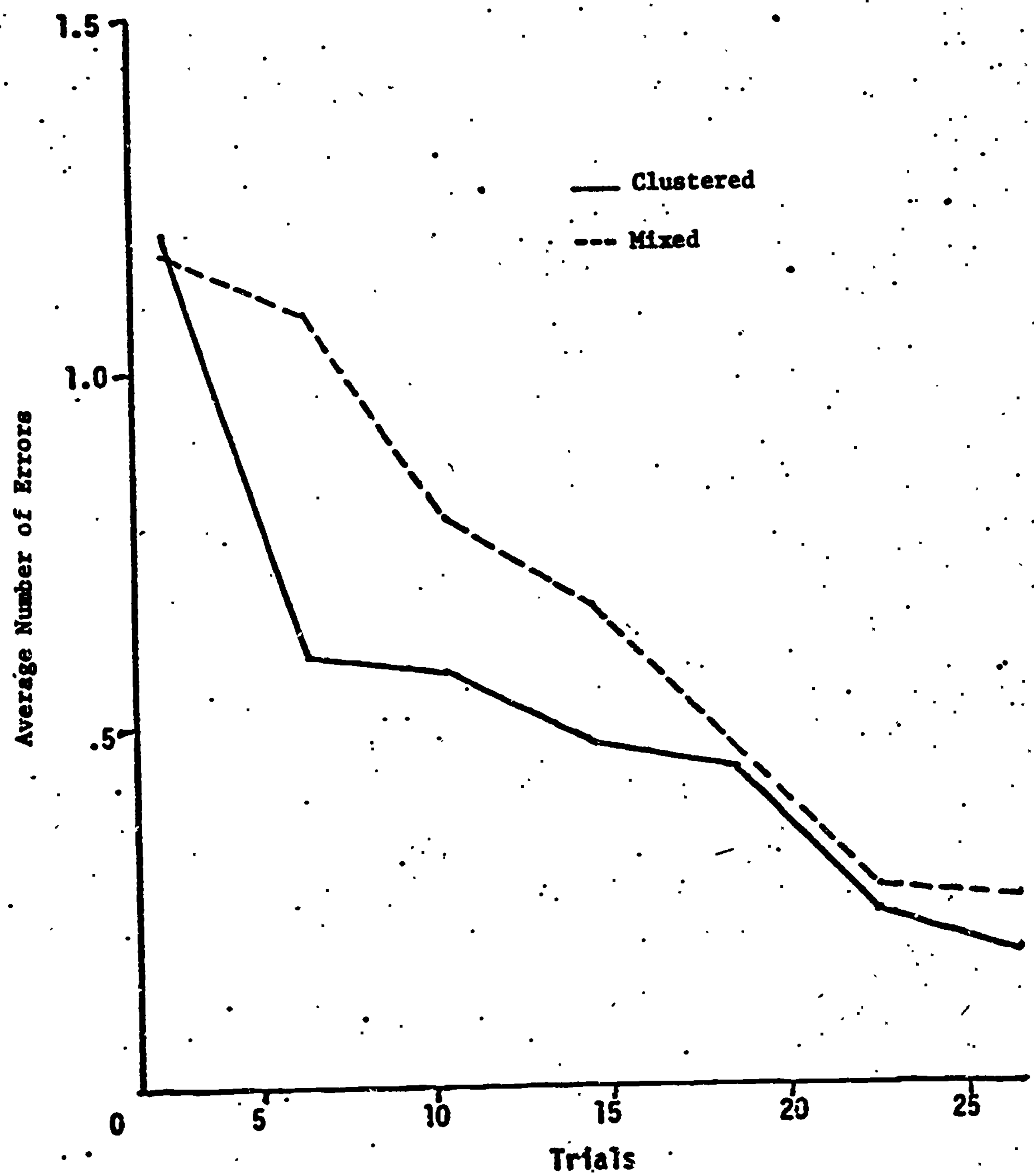

Figure 4. Rate of Learning Curves for the Mixed and Clustered Item Arrangements (Averaging Trials in Blocks of Four)

Time. Significant differences ($p < .005$ to $p < .001$) were found for nearly all comparisons of practice time among the three learning model groups. The typical mastery learning model required almost twice as much time as either of the other two models. In all cases, the adaptive mastery learning model required significantly less practice time than either of the other two models. The typical mastery learning model required significantly more practice time than the traditional non-mastery model for objectives two and four, but not for the other three objectives. This could be due to the fact that objectives two and four are fairly difficult objectives. Thus, it is more likely that the students would fail the progress checks for these objectives and spend additional time in practice. If a studnet in the typical mastery model passes the progress check on the first attempt, he receives exactly the same number of practice items as students in the traditional non-mastery model. Therefore, it is not surprising to find no significant differences in time practicing the easier objectives for these two models.

The smallest variances in practice time were obtained for the traditional non-mastery model in which all students received the same number of practice items. The largest variances in practice time were observed for the typical mastery learning model.

Thus, it can be concluded that the adaptive mastery learning model required the least amount of practice time although there was considerable variation among students in this group. It should be noted that the typical mastery learning group not only spent the most amount of time on the practice items, but also spent additional time taking the progress checks which were not required by the other models. Thus, the difference in the total amount of time required by the typical mastery learning model as opposed to the other two models is actually much greater than is reflected by comparisons of practice time only.

Amount of practice. The number of practice items was a variable for the two mastery learning models, but a constant for the traditional non-mastery model. In all cases the adaptive mastery learning model required significantly ($p < .001$) fewer practice items than the typical mastery learning model. Also, the adaptive mastery learning model required significantly fewer practice items than the traditional non-mastery learning model for all objectives except the most difficult objective. The typical mastery learning required significantly more practice items than the traditional non-mastery model for all objectives except the easiest objective. The number of items presented by the traditional non-mastery learning model was the same as the number of practice items included in the original SCIS materials, based on the authors' estimates of the amount of practice needed for each objective. From the results of this study, it appears that their estimates were suitable for the difficult objectives, but that they included more practice items than necessary for the easier objectives.

Based on the results of analyses for the total number of practice items required by each model, it was concluded that the adaptive mastery learning model is most efficient in terms of the amount of practice required. In view of the finding of no significant differences on the performance measures, this suggests that the additional practice required by the typical mastery and traditional non-mastery learning models was not beneficial.

The criterion for terminating practice for the adaptive mastery learning model was five consecutive correct answers per objective. The length of the longest string of consecutive correct answers per objective was measured for students in the other two models and compared with the constant five. With one exception the length of the longest string of consecutive correct answers per objective was significantly ($p < .0005$) greater than five for the typical mastery and traditional non-mastery models. Again, this one exception occurred for the most difficult objective under the traditional non-mastery treatment. It should be noted that there were also no significant differences in the number of practice items required for this objective between the adaptive mastery and traditional non-mastery treatment groups. If a criterion of five consecutive correct answers results in sufficient practice as was indicated by the performance results, then any more than five consecutive correct answers may be considered overpracticing. Based on the results of this study, it may be concluded that such overpracticing does not improve performance on either immediate or delayed retention tests. This is consistent with previous research findings. Further research is needed to determine if a criterion of less than five consecutive correct answers would be as efficient. Studies by Hannum (1973) and Gay (1971) suggest that as few as two consecutive correct answers may be adequate.

Aptitude. The relationships between the six aptitude measures and the variables of performance, practice time, and amount of practice were investigated with respect to Carroll's (1963) model. It was predicted that the correlation between aptitude and performance would be low under mastery learning conditions and high under non-mastery learning conditions. Quite the opposite results were found in this study in that the highest correlations between aptitude and performance were obtained for the typical mastery learning group and that the lowest correlations obtained were for the traditional non-mastery learning group.

Also, a large negative correlation was predicted between aptitude and time for mastery learning conditions and that this correlation should be near zero for traditional non-mastery learning conditions. In general, the correlations for the typical mastery and traditional non-mastery groups tend to support these predictions. However, only one of the six correlations for the adaptive mastery group was significantly different from zero. It should be noted that the time measure for Carroll's theory included all instructional and practice time for a particular topic, while the present study considered only the time spent in practice.

Carroll's theory was extended to predict a large negative correlation between aptitude and number of practice items for the mastery learning groups. Although the correlations for the two mastery learning groups were generally negative, none of them were significantly different from zero.

No single aptitude measure appeared to provide the best prediction for all conditions. In view of the extremely small sample sizes (17, 16, 18) in this study, the reported correlations may be only suggestive of the true nature of the relationships between aptitude and the various dependent variables.

Difficulty. There was a strong relationship between the difficulty of the objectives and the number of practice problems required by the adaptive mastery learning model. The relationship was not quite as strong for the typical mastery learning model. There was very little correspondence between the difficulty of the objectives and the number of practice problems established by the authors for the traditional non-mastery learning model. Thus, it was concluded that the adaptive mastery learning model provided a better adjustment of number of practice items to the relative difficulty of the objectives than either of the other two models. One advantage of the adaptive mastery learning model is that it does not require any prior knowledge of the difficulty of the objective in order to make this adjustment. It has already been mentioned that the difficulty of the objectives influenced the results obtained for several variables measured in the present study.

In summary, the results of the present study indicate that the adaptive mastery learning model produced the same high level of performance as the other two models, but required less time, fewer practice items, and minimized overpractice. In addition, the adaptive mastery learning model more readily adjusts to the difficulty of the objective. Thus, it was concluded that the use of student performance on practice items is an effective and efficient means of predicting mastery.

Distribution of Practice

Performance. There were no differences in posttest or retention test scores for the mixed and clustered item arrangements. Both groups were performing at a fairly high level. Also, there were no differences in the number of practice items required by the mastery learning models under the two conditions of item arrangements. In general, there were no differences for the number of errors made under the mixed and clustered item arrangements as measured by the ratio of correct answers to total responses.

Time. There were no significant differences in the amount of practice time required under the mixed and clustered item arrangements. However, the variance in practice time was consistently greater for the clustered item arrangement than for the mixed item arrangement. Thus, the hypothesis that the mixed item arrangement would require more practice time was not supported.

Learning rate. The only indication in the present study that there may be a difference in the effect of the mixed and clustered item arrangements was found in the two graphs of the number of errors made on the five objectives across trials. From the first graph (Figure 3) it was noted that the average number of errors dropped on every fourth trial (first, fifth, ninth, etc.). This was probably due to the fact that for each objective there were four item forms representing the four combinations of positive and negative signs for the two integers. For all objectives, the item form for every fourth trial generated a practice problem with two positive integers. These items were more familiar to the students and, therefore, easier.

The average number of errors increased for the positive-negative combinations (second, sixth, tenth trials, etc.), decreased for the negative-positive combinations (third, seventh, eleventh trials, etc.), and was inconsistent for the negative-negative combinations (fourth, eighth, twelfth trials, etc.). Such results indicate the fact that within each objective there are several different tasks requiring different rules for finding the correct answer. These results suggest that more narrowly defined tasks than the objectives employed in the present study should be used as a basis for investigating clustered versus mixed item arrangements.

For the second graph (Figure 4) the number of errors was averaged across each block of four trials. This graph more clearly illustrates a difference between the two item arrangements. The average number of errors for the two treatment groups is approximately the same on the first trial and again after the twentieth trial. However, the average number of errors for the clustered item arrangement initially decreases more rapidly and then levels off.

Thus, no conclusions can be drawn from the present study regarding the effectiveness of mixed and clustered item arrangements. However, Figure 4 indicates the possibility of a difference between the two treatments. Further research employing more narrowly defined tasks should be conducted.

Future Research

A number of different research studies are suggested by the results of this study. Various criteria for terminating practice in a computer-based adaptive mastery learning model should be compared. This study employed a criterion of five consecutive correct answers per objective to predict mastery. Other studies (Hannum, 1973; Gay, 1971) indicate that as few as two consecutive correct answers may be sufficient. Does the number of consecutive correct answers needed to predict mastery vary according to student or task characteristics? This study should be replicated using different subject matter. In addition, the adaptive mastery learning model should be employed with a larger sample size in order to more adequately investigate the relationships between aptitude and various other measures.

Several studies could also be conducted regarding the mixed and clustered item arrangements. Research employing more narrowly defined tasks should be conducted. In other words, a set of items for a task should all involve the use of the same rule or algorithm for their solution. In addition, the effectiveness of the two item arrangements should be investigated under both immediate and delayed practice conditions.

References

Bassler, O. C., Curry, D., Hall, W., & Mealy, E. An investigation of two instructional variables in learning non-metric geometry. School Science and Mathematics, 1971, 71, 441-450.

Block, J. H. Introduction to mastery learning: Theory and practice. In J. H. Block (Ed.), Mastery learning: Theory and practice. New York: Holt, Rinehart, and Winston, 1971. (a)

Block, J. H. Mastery learning in the classroom: An overview of recent research. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, February 28, 1973.

Bloom, B. S. Recent developments in mastery learning. Educational Psychologist, 1973, 10(2), 53-57.

Carroll, J. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Crothers, E. J. Presentation orders for items from different categories. Technical Report No. 71. Psychology Series. Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965.

French, J. W., Ekstrom, R. B., & Price, L. A. Manual for kit of reference tests for cognitive factors. Princeton, N. J.: Educational Testing Service, 1963.

Gagne, R. M., Mayor, J. R., Garstens, H. L., & Paradise, N. E. Factors in acquiring knowledge of a mathematical task. Psychological Monographs, 1962, 76(7, Whole No. 526).

Gay, L. An investigation into the differential effectiveness for males and females, of three CAI treatments on delayed retention of mathematical concepts. Tech Memo No. 12. Tallahassee, Fla.: CAI Center, Florida State University, 1969.

Gay, L. Temporal position of reviews and its effect on the retention of mathematical rules. Tech Report No. 18. Tallahassee, Fla.: CAI Center, Florida State University, 1971.

Gibson, J. R. Transfer effects of practice variety in principle learning. Unpublished doctoral dissertation, University of California, Berkeley, 1969.

Hannum, W. H. A study of selected factors influencing the retention of rules. Unpublished doctoral dissertation, Florida State University, Tallahassee, 1973.

Johnson, K. A. The distribution of topics within programs. NSPI Journal, 1970, 9(1), 14-15,18.

Millman, J. Determining test length: Passing scores and test lengths for objective-based tests. Unpublished manuscript, Instructional Objectives Exchange, Los Angeles, California, 1972.

Reynolds, J. H., & Glaser, R. Effects of linear and spiral programming upon amount and variability of learning. In J. P. DeCecco (Ed.) Educational technology: Readings in programmed instruction. New York: Holt, Rinehart and Winston, 1964.

Rivers, L. Development and assessment of an adaptive strategy utilizing regression analysis techniques for the presentation of instruction via computer. Tech Report 27. Tallahassee, Fla.: CAI Center, Florida State University, 1972.